

日本史史料の長期利用とデータ共有・連結化に向けた システム環境整備

渋谷 綾子・山田 太造・渡邊 要一郎・平澤 加奈子（東京大学 史料編纂所）

大向 一輝（東京大学大学院人文社会系研究科）

金子 拓・山家 浩樹・保谷 徹（東京大学 史料編纂所）

東京大学史料編纂所で実施する「人文学・社会科学データインフラストラクチャー構築推進事業」を紹介する。本事業は、人文学・社会科学研究に係るデータを分野や国を超えて共有・活用する総合的なシステムの構築を目的とする。史料編纂所では、国内外へのデータの公開・発信と海外機関との連携強化、データベース英訳事業等に取り組んでおり、大学や図書館・博物館等に所蔵される史料のデータ収集と共有化を進め、人文学分野におけるデータインフラ構築の拠点としての活動を進めている。本論文は、日本史史料の長期利用をみすえた情報環境の整備と手法、今後の課題、特に人文学と社会科学の接続に関する課題について述べる。

Developing system environment for long-term utilisation, data sharing, and linking of Japanese historical materials

Ayako Shibutani / Taizo Yamada / Yoichiro Watanabe / Kanako Hirasawa (Historiographical Institute, The University of Tokyo) / Ikki Ohmukai (Graduate School of Humanities and Sociology, The University of Tokyo) / Hiraku Kaneko / Koki Yanbe / Toru Hoya (Historiographical Institute, The University of Tokyo)

This paper deals with the ‘Program for Constructing Data Infrastructure for the Humanities and Social Sciences’ by the Historiographical Institute (HI) of the University of Tokyo. This project seeks to construct an integrated system for sharing and utilising data in the humanities and social sciences regardless of academic field or country. The HI distributes data inside and outside Japan, promoting collaborations with other countries, and is developing an English database of translations. We discuss the construction of an information infrastructure and its applications for long-term utilisation, as well as future issues regarding the links between the humanities and social sciences.

1. はじめに

東京大学史料編纂所（以下、史料編纂所、HI）は、明治以来 150 年にわたって国内外の日本史史料を調査・研究し、日本史の基幹史料集を編纂・出版することを研究所のミッションとしてきた。1980 年代からいち早く情報機器の導入につとめ、1984 年より史料編纂所歴史情報処理システム（SHIPS : Shiryohensan-jo Historical Information Processing System）の構築を開始した。SHIPS には、これまでに史料編纂所で作成してきた目録・画像・本文・人名・地名・文字などのデータが蓄積されている。1997 年には構築した各種データベースのウェブ公開を開始、最先端の歴史情報学研究を切り拓いた。2006 年、歴史情報研究の中核組織として前近代日本史情報国際センター（以下、情報センター）が設けられた。5 つの研究部門や附属画像史料解析センターを基盤として、文部科学省から共同利用・共同研究拠点として認定され、

日本史史料の研究資源化研究拠点の活動を行っている。

史料編纂所は、国内・海外の日本史史料を調査して複製による史料収集を行うと同時に、長期にわたって蓄積した貴重な原本史料を多く所蔵している。かかる収集史料や所蔵史料をデジタル化し、史料解析の成果をメタデータとして付与し公開している。また、長年の研究蓄積を SHIPS データベースと呼ぶデータベース検索サービスより発信しており、古代から明治維新までの日本史の史料情報・歴史情報を、史料目録データベース・史料全文データベース・史料画像データベース、あるいはくずし字解読字典・歴史絵引きなどのツール型データベースなど、40 件ほどの各種データベース（データ件数 680 万以上）に搭載している。加えて 2,000 万件の史料画像を有し、人文系では屈指の日本史データコレクションとなるまでに成長した。史料の調査・収集や史料データの作成は、130 年を超える編纂所の編纂・研究事業に基づくものである。

2019年度からは、「人文学・社会科学データインフラストラクチャー構築推進事業」（以下、データインフラ事業）[1]の拠点機関として、人文学を対象とした研究データの共有、および利活用を促進するための情報環境の整備を進めている。データインフラ事業は、人文学・社会科学に属するデータを分野や国を超えて共有・利活用する総合的なシステムの構築を目的とする。史料編纂所では、国内外へのデータの公開・発信と海外機関との連携強化、データベース英訳事業等に取り組んでおり、大学や図書館・博物館等に所蔵される史資料のデータ収集と共有化を進め、人文学分野における拠点機関としての活動を進めている。

本論文は、事業の取り組みを報告するとともに、どのようなシステムの構築を行い、それが人文学学術情報基盤へどのような影響をもたらすのか、類似インフラとの差異やメタデータ等の相互運用性、また人文学と社会科学の接続に関する課題解決について述べる。

2. 人文学・社会科学データインフラストラクチャー構築推進事業

情報環境は、情報システムあるいはシステムに格納されるデータだけではなく、情報を収集し、管理するための一連のワークフロー、第三者が利用する際の規約、国内外の研究コミュニティとの連携などを含む広汎な概念である。データインフラ事業ではこれらの課題について、(1) データアーカイブ機能の強化を中心とした「共有化」、(2) 海外発信・連携機能の強化に代表される「国際化」、

(3) 他機関のデータとの連携を目的とする「連結化」、という3つのユニットを設けてそれぞれ対応している。2020年10月までの成果は次の通りである。

2.1. データアーカイブ機能の強化（共有化）

史料編纂所が保有する大規模データベース（Hi-CAT [2]、Hi-CAT Plus [3]（図1））の機能強化、データの拡充を通じて、日本史研究のみならず、人文学研究コミュニティへの貢献を行ってきた。機能面では、データを長期保存・長期利用するためのシステム環境整備として、史料編纂所の史料調査活動全体を管理する「探訪進捗管理システム」を、OAIS（Open Archival Information System）参照モデル[4]にもとづき、史料データの保存に関する記述を整備した（図2）。

史料編纂所では、国内外の日本史史料の調査・収集を「史料探訪（または探訪）」と呼ぶ。図3のように、探訪に関する計画を立てた後、史料編纂所図書室に対してそれを申請する。探訪の申請では、探訪コード申請者、史料（群）名、史料（群）名ヨミ、原蔵者、原蔵者ヨミ、原蔵者住所、探訪名称、撮影代表者、撮影メンバー、撮影年月日、

備考などを記入する。このうち、史料（群）名、撮影代表者は必須項目である。これを行うと探訪コードが発給される。発給は史料編纂所図書室にて行われる。探訪を行った後に、先の申請内容の確認を経て、図書室へ画像を受け渡し、アクセス権限についても連絡する。これに応じて画像の整備、データベースへの登録、画像サーバへの登録、画像のアクセス権限設定が行われる。史料請求の場合においてもこの探訪コードを用いて行われる。このように探訪コードはあらゆる場面での管理コードであると位置づけることができ、探訪を行う研究者、探訪コードの管理を行う図書室、史料の利用者などにも利用されていく。

図1 Hi-CAT Plus の検索画面
Figure 1 Search Screen of Hi-CAT Plus.

作成された目録と画像は、SHIPS におけるデータベース検索サービスである SHIPS データベースに格納される。史料編纂所所蔵史料の場合は、SHIPS データベースの1つである所蔵史料目録データベース（Hi-CAT）、史料編纂所以外の所蔵史料は Hi-CAT Plus へ登録される[5]。

史料編纂所では年間 50 ヶ所以上の探訪を行い、1度の探訪で何箇所もの機関等へ訪れることから、年間 200 以上の探訪コードが発給される。探訪した史料画像がデータベースへの登録へ登録されるまでに1年以上かかることもあり、また各探訪の状態は 16 段階に分類されるため、探訪という行為を管理するシステムが求められた。そこで「探訪進捗管理システム」という探訪の進捗を管理するためのシステムを構築し、2012 年より運用を開始した[5]。2020 年 10 月時点で 3,248 件の史料探訪進捗に関するデータが管理されている。運用後 10 年を経過し、少しずつ運用面での変更が求められている。例としては、デジタル化や撮影に関わる機材、データ利用条件である。これらについて、データ項目の精査やデータ公開までの

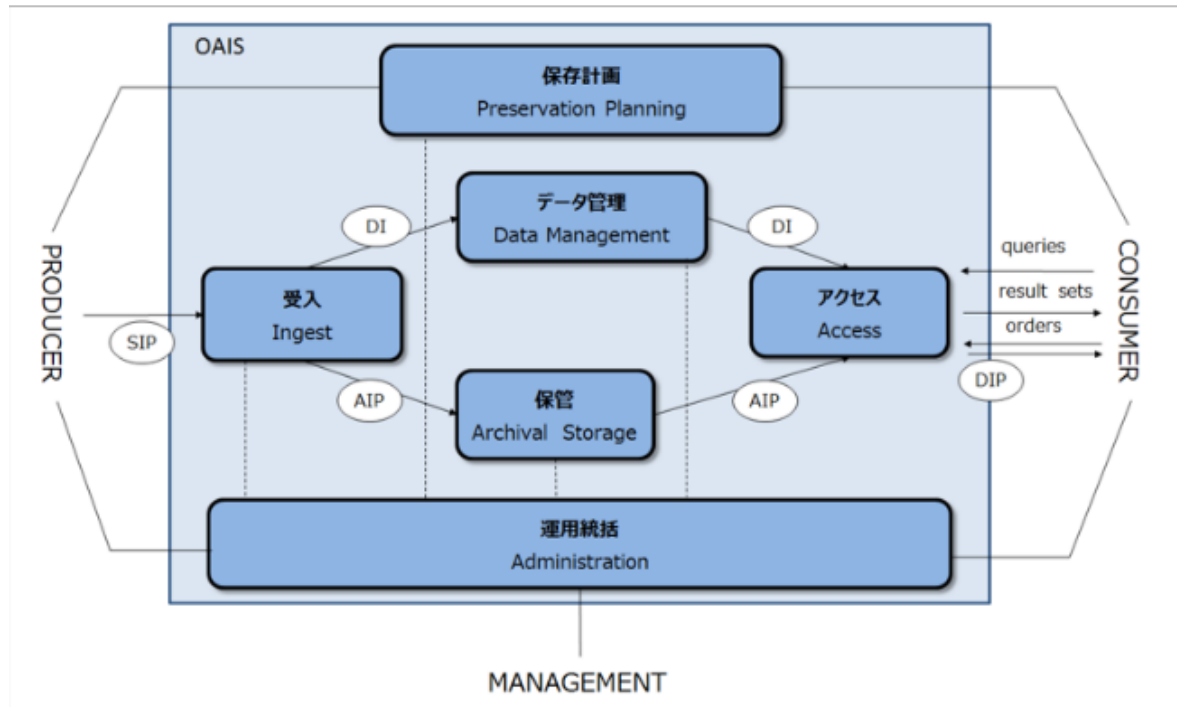


図2 史資料データの長期保存・長期利用システム
Figure 2 Long-term Preservation and Utilisation of Historical Materials.

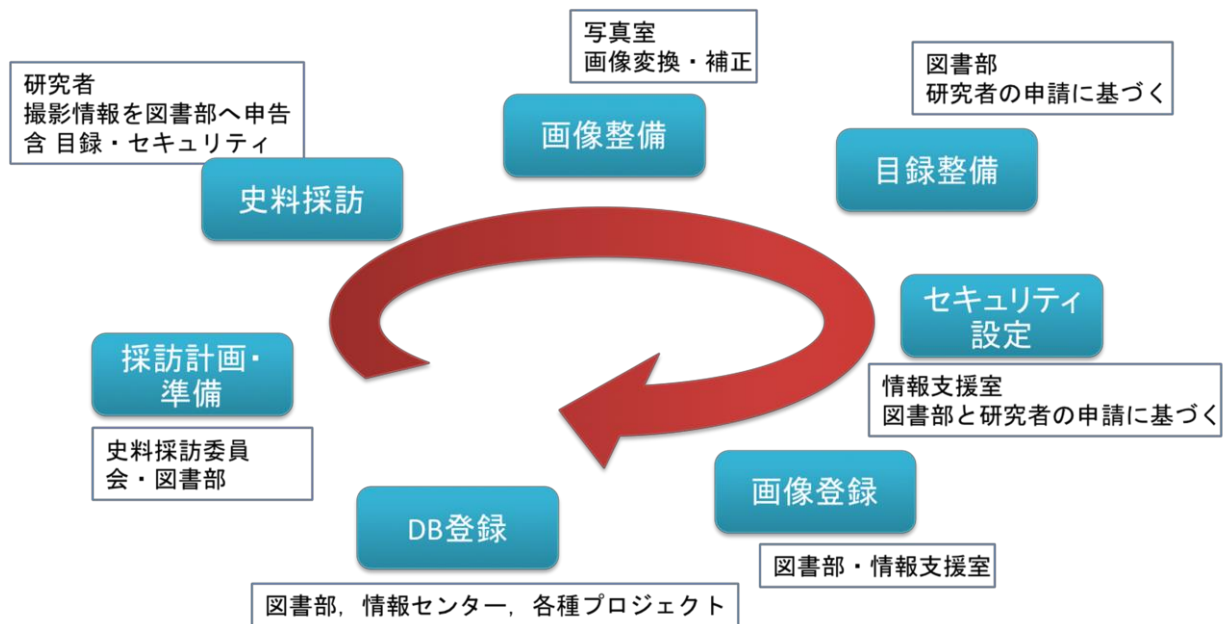


図3 史料探訪の流れ
Figure 3 Flow of material investigation and collection

フローの見直し, さらには史料探訪に関わる各種マニュアル等の見直しを開始しており, システム等への具体的な適用を行っている。

2. 2. 海外発信・連携機能の強化（国際化）

1930 年頃より, 史料編纂所は海外に所在する日本関係史料の調査事業を開始し, 日本学士院・ユネスコ・国際学士院連合等の支援を得て, マイクロフィルムによる史料収集を開始した。また,

これまで不十分であったロシア・中国方面の史料フローの見直し、さらには史料採訪に関わる各種マニュアル等を見直しを開始しており、システム等への具体的な適用を行っている。

フローの見直し、さらには史料採訪に関わる各種マニュアル等を見直しを開始しており、システム等への具体的な適用を行っている。

2. 2. 海外発信・連携機能の強化（国際化）

1930 年頃より、史料編纂所は海外に所在する日本関係史料の調査事業を開始し、日本学士院・ユネスコ・国際学士院連合等の支援を得て、マイクロフィルムによる史料収集を開始した。また、これまで不十分であったロシア・中国方面の史料フローの見直し、さらには史料採訪に関わる各種マニュアル等を見直しを開始しており、システム等への具体的な適用を行っている。

収集も系統的に実施している。中でも、ロシア国立歴史文書館、ロシア国立海軍文書館からはデジタルデータの提供を受けており、共有方法について研究会を開催し、公開を進めている。さらに、横浜開港資料館と協議の上、英国 The National Archives が所蔵する英国外務省日本関係史料（FO46・FO26）のデジタル画像約 64 万コマを、Hi-CAT Plus を介して公開している。

史料編纂所 SHIPS データベースの 1 つである『維新史料綱要』データベース英語版構築に取り組み、2019 年度に試験版を作成し、2020 年度中の公開を予定している。これは、2019 年度から進めている「維新史料研究国際ハブ拠点形成プロジェクト」（維新ハブ PJ）と連携している。この維新ハブ PJ は、幕末維新史データベースである維新史料綱要データベースの英語化と、そのための歴史用語・史料用語の英訳グロッサリー研究を行うものである。プロジェクトとデータインフラ事業との連携としては、特に『維新史料綱要』データベースにおけるユーザインタフェースのみならず、網文の英訳化も行い、史料用語・歴史用語の英訳グロッサリー作成作業を推進している。2020 年 10 月末段階で全 10 巻中 3 巻前半約 6,000 件の網文を英語化し（図 4）、逐次登録を進めている。これに関連して、2019 年 12 月には国際研究集会「維新史料研究と国際発信」[6]を開催した。この研究集会では、海外研究者の視点から意見を得るとともに、プロジェクトや史料編纂所所蔵幕末維新史料、幕末維新史データベースについての国際発信を行った。

人文情報学部門による SAT（大正新脩大藏經テキストデータベース）では、英語版インターフェースのみならず、英語翻訳など海外研究者の利用を見越した先進的な整備を具現化している。これを SHIPS へ適応化させる可能性の検討を進め、人文学研究における共通化の可能性を検証した。さらに、ベトナム・ノム保存財団からは、データ

収集という観点でデジタルデータの提供を受けている。

No.	Details	Date (Japanese calendar)	Kōyō Entry
1	Details	弘化3年4月15日	Lord of Kanazawa domain Maeda Nariyasu (Kaga no kami) goes up to Edo castle to receive leave to return to his domain. Lord of Kagetsu domain Shimazu Narioki (Osumi no kami), lord of Aizu domain Matsudaira Katataka (Higo no kami), lord of Kubota domain Satake Yoshiatsu (Ukyō no daibu), lord of Nihonmatsu domain Tanabe Nagatomi (Sakyo tairō) lord of Kōchi domain Yamauchi Toyotaru (Tosa no kami), heir to the lord of Kanazawa domain Maeda Yoshiyasu (Chikuzen no kami), lord of Yonezawa domain Uesugi Narioki (Danjōdaijū), lord of Okayama domain Ikeda Yoshimasa (Kura no kami), lord of Uwajima domain Date Munenari (Tōtomi no kami), lord of Nakamura domain Sōma Dazemon no sukei, lord of Nagato Fuchū domain Mori Motoyuki (Sakyo no suke), lord of Izumi domain Sengoku Hisatoki (Sanuki no kami), lord of Hōryō domain Rokugo Masatsune (Hyōgo no kami), lord of Saeki domain Mori Takayasu (Awa no kami), lord of Hachinohe domain Hasegawa Nobuyori (Tōtomi no kami), lord of Ichinohe domain Tamura Kunyuki (Ukyō no daibu), lord of Ōmori domain Wakebe Mitsusada (Wakasa no kami), lord of Nishijō domain Kihashi Nagakazu (Shimoda no kami), lord of Komono domain Hijikata Katsuyoshi (Bitchū no kami), lord of Yanagimoto domain Oda Hideaki (later Nobuaki, Aki no kami), lord of Tadotsu domain Kyōgoku Takataru (Ki no kami), and lord of Shimodate domain Tachibana Tanemaru (Shuzen no shō) go up to Edo castle for audience ceremonies in conjunction with their arrival in the city.
2	Details	弘化3年4月18日	Lord of Tsuyama domain Matsudaira Naritami (Mikawa no kami), lord of Fukui domain Matsudaira Yoshinaga (Echizen no kami), lord of Tokushima domain Hachisuka Narihiro (Awa no kami), lord of Kuramoto domain Hosokawa Narimori (Etchū no kami), lord of Sendai domain Date Yoshikuni (Mutsu no kami), lord of Hagi domain Mori Yoshichika (Daisenryū), lord of Tottori domain Ikeda Yoshiyuki (Izumi no kami), lord of Otsu domain Maeda Yoshihira (Bingo no kami), lord of Ōzu domain Katō Yasumoto (Tōtōmi no kami), lord of Sando domain Kuki Takanori (Nagato no kami), lord of Sadowara domain Shimazu Tadahiro (Awaji no kami), lord of Hiji domain Kinoshita Toshiatsu (Saemomotsu), lord of Akō domain Mori Tadanori (Etchū no kami), lord of Naegi domain Toyama Tomosaki (Mino no kami), lord of Ōg domain Nakatsuna Tadashika (Kō no kami), lord of Tottori domain Isen domain (later, Shikano domain) Ikeda Nakamori (Ki no kami), lord of Uto domain Hosokawa Yūhisa (Bizen no kami), lord of Iyo Yoshida domain Date Munetaka (Wakasa no kami), lord of Sonobe domain Koide Fusanori (Shinano no kami), lord of Okayama domain branch domain (later, Kamogata) Ikeda Masayoshi (Shinano no kami), lord of Kashiwabara domain Oda Nobusada (Izumo no kami), lord of Nimi domain Seki Nagamichi (Tajima no kami), lord of Toyooka domain Kyōgoku Takayuki (Kai no kami), lord of Mori domain Kurashima Michihiro (Iyo no kami).

図 4 『維新史料綱要』データベースの英語版

Figure 4 English Version of Summary Database of the Ishin Shiryō.

2. 3. 他機関のデータとの連携（連結化）

コロナ禍のもとでデータへのアクセスは急増しており、SHIPS データベースにおける画像データへのアクセス件数は 2019 年度では約 850 万件だったが、2020 年度は過去最大の 1,000 万件に達する見込みである。アクセス件数の増加要因の 1 つが、COVID-19 の流行に伴う史料原本へのアクセス制限であることは明白である。コロナ禍では一層のデジタル史料画像への重要性は増していくと考えられるため、それへのアクセス方法および利用条件の整備は、もはや公開者における必須事項と言っても過言ではない。

史料編纂所では、史料画像へのアクセス手段として、IIIF（International Image Interoperability Framework）[7]を利用した取り組みを行ってきた。具体的には、2015 年 12 月に公開したウェブ画像ビューア SHIPS Image Viewer において Image API を利用しており、2020 年 3 月にウェブ公開した Hi-CAT Plus にて IIIF Presentation API を用いた画像ビューアを導入した。

史料編纂所所蔵史料の画像データの利用条件は 2019 年 4 月に公開した。これを図 5 に示す。この利用条件は、東京大学デジタルアーカイブズ構築事業のもと、デジタル化を行った備後福山藩阿部家史料 55,000 件についても適用している。他機関所蔵史料画像データについても、利用条件を設定している。例えば、宮内庁書陵部が所蔵する天皇家・公家文庫史料群等の画像データ 56 万件余りについては、所蔵機関との調整のもと、利

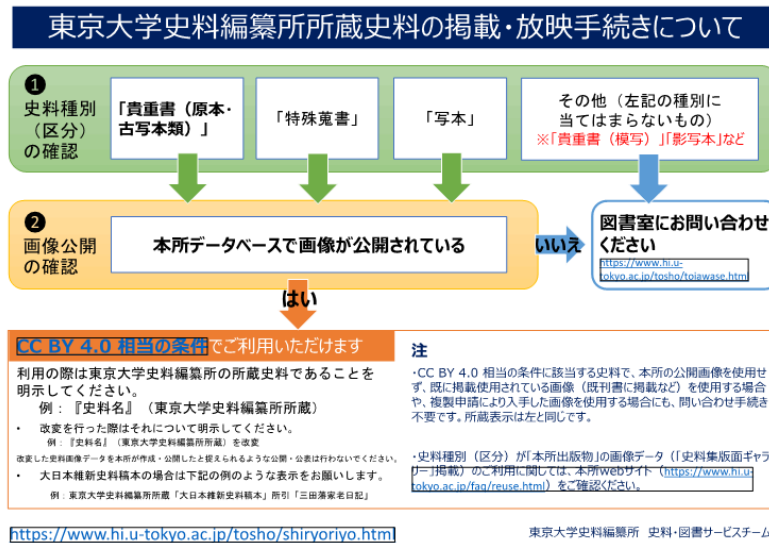


図5 所蔵史料の利用条件
Figure 5 Terms of Use of Image Data at HI.

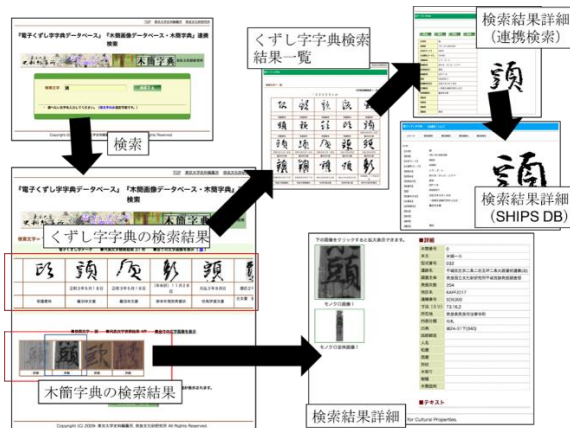
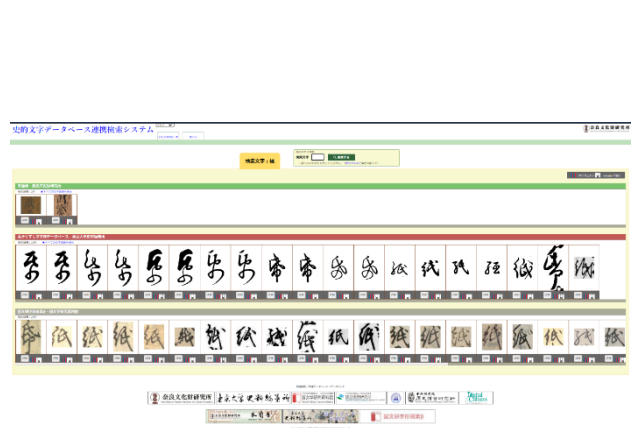


Figure 6 Multi-database Searching of “Wooden Tablet Database” and “Database of Kuzushi Ji.”



用条件の整備を行い、2020年3月にHi-CAT Plusを介し公開した。

2009年10月に奈良文化財研究所と史料編纂所は「『電子くずし字字典データベース』『木簡画像データベース・木簡字典』連携検索」（図6）を公開し、文字画像の連携検索システムを実現した。史料編纂所からは、283,500件の文字画像を公開している。本システムをさらに改善するため、機械可読である検索結果返戻を実現するAPIの整備、二次利用可能な文字画像データの提供、文字データの増強に取り組み、新しい文字画像連携検索システム「史的文字データベース連携検索システム」（図7）を2020年3月にベータ版を公開し、2020年10月には本運用を開始した[8]。

データ量だけでなく、字種・文字の年代や地域といった文字データの「拡大」を実現するため、

国文学研究資料館、国立国語研究所、京都大学人文科学研究所、漢字字体規範史データセット保存会、台湾中央研究院と連携した。さらに、機械可読可能なデータの共有を図るため、RDFによるデータをやり取り可能なREST型のAPIを整備した。画像データはIIIF Presentation APIに対応させており、オープンデータ（CC BY-SA 相当以上）としてデータ利用条件を設定している。

Hi-CATの画像ビューアは、データ利用条件が提示されていない。これを可能にするために、Hi-CAT Plusでの成果をHi-CATへ適用する準備を進めている。あわせて、2019年より運用を開始した史料集版面ギャラリー（史料編纂所が編纂・出版した史料集の版面画像ギャラリー）[9]への適用も進めており、特定のデータベースを紹介すること

なく、研究素材となるデータを提供できるように、データセットとして整備し、提供する予定である。

3. 人文学と社会科学データの横断検索システムの構築と関連する課題

3. 1. 人文学・社会科学総合的データカタログ

データインフラ事業では、日本の人文学・社会科学研究機関が公開する研究データの情報を収集する「人文学・社会科学総合的データカタログ (JDCat : Japan Data Catalogue for the Humanities and Social Sciences)」(図 8) の構築を進めている。JDCat は、拠点機関が収集・整理・提供するデータを横断的に一括検索できるものである。社会科学拠点、人文学拠点のデータベースならびにデータリポジトリからのメタデータの自動収集機能が設けられており、格納されたデータは新たに構築される研究データ検索サービス CiNii Research や海外のデータ検索サイトに提供される予定である。

JDCat の基本方針は、社会科学分野および人文学分野の拠点機関が保有するデータの相互運用性を高め、長期にわたり使用できるように設計することである。さらに、ユーザビリティを高めるため、国内外の人文学・社会科学を中心とした研究者に必要な情報が過不足なく記述され、さまざまな分野のデータ登録者(データキュレーターやデータ提供者)にとって理解や登録が容易な形式であることが求められる。そこで、本事業では上記の要求を満たすべく、JDCat メタデータスキーマと呼ばれる仕様の策定が進められている。

マと呼ばれる仕様の策定が進められている。

JDCat メタデータスキーマは、社会科学分野で広く用いられる国際標準である DDI (Data Documentation Initiative)、機関リポジトリのメタデータ交換に用いられる JPCOAR (オープンアクセスリポジトリ推進協会) スキーマ[10]をベースとして設計されているが、人文学分野全般に適用できるように整備する必要があった。

2020 年現在まで、史料編纂所では史料データ(目録や画像)の蓄積・提供・分析を行う中で、関連する研究情報(たとえば顕微鏡撮影画像や各種の解析結果など)を紐付ける形で収集を進めてきた。この知見にもとづき、JDCat メタデータスキーマの各項目の解釈を明確にしたほか、人文学データに特有の文字に関する情報の追加のほか、曖昧さを含む地理的範囲・時間的範囲の記述方法を検討した。

3. 2. 統制語彙の策定：人文学と社会科学の接続における課題

情報検索において、索引語として利用する語を限定し、その意味範囲や使用方法を規定する統制語彙が必要となる。JDCat の詳細検索では、ファセットとしてあらかじめ定められた項目による絞り込み検索と、それ以外の項目を対象とした全文検索の 2 種類の方法が採られる。前者で用いる統制語彙(CV : Controlled Vocabulary)には、史料編纂所のデータや今後データの登録が予想される人文学分野全般を網羅すること、国内外の利用を促進するため、国際的に広く利用されている

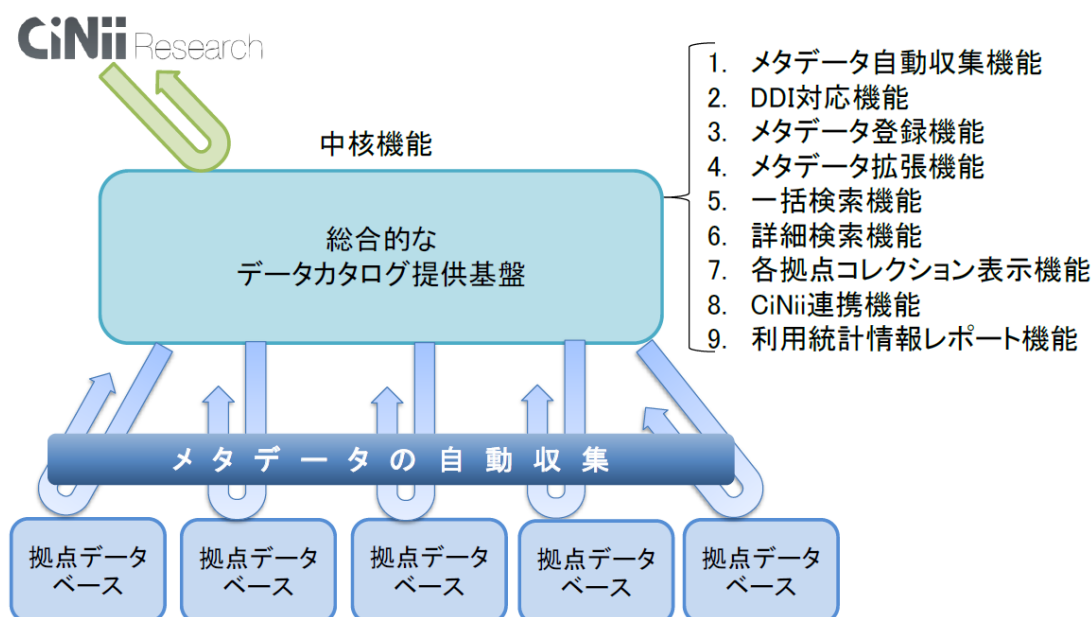


図 8 「人文学・社会科学総合的データカタログ (JDCat)」のイメージ[11]

Figure 8 Image of “Japan Data Catalogue for the Humanities and Social Sciences (JDCat)” [11].

ものを用いること、検索精度を向上させるため、必要な語彙の網羅を優先すること、といった要求がある。本事業では、データの意味内容を示すトピック CV と、データの種類を表すデータタイプ CV の検討を行った。

既存のトピック CV は、人文学全般を扱うもの、かつ国際的に利用されているものはなく、日本やアジア系の語彙が不十分であり、粒度にもムラがある。そこで、幅広い分野の資料を扱う図書館の分類を使用することとした。中でも、日本の史資料に関する分類項目を含む日本十進分類法 (NDC) を採用した。

トピック CV では、NDC 第 10 版の第 2 次区分表と第 3 次区分表を合わせることによって、芸術・美術、文学は詳細な語彙を使用可能とし、データタイプや対象時期等の項目を組み合わせた詳細な検索の実現が可能となった。

データタイプ CV は当初、社会科学で用いられる内容とデータ形式の組み合わせが提示され、人文学のデータタイプ CV の策定にあたっては、社会科学とのバランスが求められた。ここで、人文学と社会科学との接続における課題の 1 つが生じた。

データタイプ CV の候補として、史料編纂所 Hi-CAT のカテゴリ、HASSET Information Materials, JPCOAR スキーマの資源タイプが挙げられた。しかし、これらの語彙はトピック CV の内容を一部含んでいるとともに、先行して定義された社会科学のデータタイプ CV とは概念の粒度が大きく異なっていた。そこで、人文学では相互運用性を重視し、新規の語彙を設計した。

2020 年現在確定したデータタイプ CV は、社会科学では量的調査 (quantitative research)、公的統計 (official statistics)、質的調査 (qualitative research) を大項目とし、人文学では、史資料 (materials)、出版物 (publication)、事項 (fact)、データベース (database)、その他 (other) を大項目とし、合計 31 のエレメントから構成される。今後は、史料編纂所のデータを JDCat メタデータスキーマで記述し、データカタログへの登録を進める。

3. 3. JDCat と既存の関連システムとの差異

データインフラ事業で構築する JDCat と関連する既存のシステムとしては、nihuINT (nihu INTegrated retrieval system, 高度連携・統合検索システム) [12]、ジャパンサーチ[13]、Europeana [14] がある。

nihuINT は人文科学研究資源を統合するプラットフォームであり、人間文化研究機構 6 機関、および国会図書館・京都大学地域研究統合情報センターなどのデータベースを統合的に検索可能な

ものである。ジャパンサーチは、書籍等分野、文化財分野、メディア芸術分野など、さまざまな分野のデジタルアーカイブと連携し、多様なコンテンツのメタデータをまとめて検索できるシステム「国の分野横断型統合ポータル」である。主として美術分野への強みをもつ。Europeana は文化遺産のための EU のデジタルプラットフォームである。参加機関は 3000 以上にのぼり、アムステルダム国立美術館や英国図書館、ルーブル美術館など国際的に著名なところから、欧州連合のすべての加盟国における地域の文書館や美術館・博物館までが含まれている。

これらのシステムと JDCat は、人文学・社会科学のさまざまな研究データの情報を横断的に一括検索できる点で大きく異なっており、先駆的である。JDCat のデータ利用者は、自分の研究テーマに合ったデータを横断検索可能であり、データ提供者の側は研究プロジェクトのリポジトリを、機関リポジトリを利用して運用可能であるというメリットが想定できる。

さらに、JDCat とあわせてデータインフラ事業では、分野横断的なオンライン分析システムの開発を進めている。各拠点機関が整備したデータの中での特に重要なものを対象に、オンラインで統計分析を行い、これらの分析プログラムや結果を分析ノートとして利用者が設定したグループで共有することを目的としたシステムである。拠点機関によっては、データの保護のためにアクセス制限を設け、利用申請制度や利用期間がある。オンライン分析システムは、それぞれの拠点機関の状況に応じた設計が進められている[15]。史料編纂所では、人文学でのデータカタログの形式やオンライン分析手法の検証を行い、プロトタイプを重ねながら深化させている。

3. 4. コロナ禍における人文学研究の推進

COVID-19 流行は、これまでの人文学研究の基礎とも言うべき史料原本へのアクセスへの制限をもたらした。この人文学研究推進の危機的状況の中、デジタル画像データ提供は、この状況を打破するための最有力手法であると認知され、ウェブを介した発信力強化が一層求められている。

その一方で、人文学分野においては、データ管理・利活用がそれほど進んでいない状況にある。背景として、基本的に人文系分野の研究が研究者個人の営為によって完結してしまい、著作のかたちで研究成果が示されると、その研究過程で調査・集積したデータが検証可能なかたちで共有されないことが考えられる。たとえば、MLA (博物館・図書館・文書館) での所蔵資料のデータ公開が急速に進む一方、研究過程で集積した学術資料をどのようにデータ化し、共有化していくのかというプロセスは、いまだ確立していない。

史料編纂所では、個人やチームの調査成果は独自開発した採訪進捗管理システムにより、すべて研究所の資源として登録・共有されて公開利用に供される原則として決定し、このシステムはすでに10年におよぶ運用実績がある。集積した史料情報や研究成果を散逸させてしまうことなく、データとして保存し、研究組織としての蓄積を世界へ公開・発信する役割を果たしてきた。この組織的取り組みこそが、史料編纂所の最大の強みである。この実績と経験は、MLAや各種の資料を扱う人文系の研究機関に応用可能なものであり、その実践こそがデータインフラ事業の人文学拠点としての最大の成果である。これらの成果は、コロナ禍のもとでの人文学研究の推進の一助になると考える。

4. おわりに

史料編纂所は、日本古代から明治維新期に至る前近代日本史研究に欠かせない史料の収集・編纂・出版を150年に涉って実施してきた。この研究事業は他に例を見ないことから独自性が非常に高いが、それだけではなく、所属する研究者個人々人というよりも、史料編纂所組織として体系的に取り組んできたことから、計算機が登場する前から広く日本史史料や研究成果を共有・管理してきたという特徴がある。

史料編纂所による史料データ提供は、日本史学研究を支えるだけでなく、人文学研究やさらには情報学・地震学との連携に見られるように、いまや他の研究領域にも実施されている。また、他の分野の研究成果によって、日本史研究の推進にもつながっている。他機関とのシステム連携を行うための体制づくりも進めており、令和3年度から5年間実施する「データ駆動型歴史情報研究基盤の構築」事業など、人文学を対象とした新たなプロジェクトの基盤整備も確立させた。

このような背景をもって、データインフラ事業では、日本史学や仏教学を中心とした人文学データの共有基盤の強化を目指しており、国内外の人文学研究者・人文学に関係する研究機関だけでなく、広く人文学データを研究資源として永続して提供するためのシステム運用を進めている。これまでに実施した内容は、史料編纂所そのものの存在意義と同じくするものであり、事業の終了後も継続する予定である。

謝辞

本論文は、JSPS 人文学・社会科学データインフラストラクチャー構築推進事業の成果の一部である。関係者・関係機関に深く感謝申し上げたい。

参考文献

- [1] “人文学・社会科学データインフラストラクチャー構築推進事業”. <https://www.jsps.go.jp/j-di/index.html>, (参照 2020-09-03).
- [2] 加藤友康：研究成果報告書「WWW サーバによる日本史データベースのマルチメディア化と公開に関する研究」(1999). <http://www.hi.u-tokyo.ac.jp/personal/kato/index.htm> (参照 2019-10-23).
- [3] 大内英範, 山田太造, 高橋典幸, 綱川歩美, 林譲, 保谷徹, 山家浩樹, 横山伊徳. Hi-CAT Plus: デジタル史料の検索・閲覧システム, じんもんこん 2011 論文集, vol.2011, no.8, pp.105-110 (2011).
- [4] “The Reference Model for an Open Archival Information System”. <http://www.oais.info>, (参照 2020-09-03).
- [5] 山田太造, 井上聡, 山家浩樹. 日本史史料データ流通基盤に向けた歴史データリポジトリの整備, じんもんこん 2019 論文集, vol.2019, pp.3-10 (2019).
- [6] “東京大学史料編纂所国際研究集会「維新史料研究と国際発信」”. <http://www.hi.u-tokyo.ac.jp/news/2019/ishinsiryoudenkyuu201912.pdf>, (参照 2020-11-04).
- [7] “International Image Interoperability Framework”. <https://iiif.io>, (参照 2020-09-03).
- [8] “史的文字データベース連携検索システム”. <https://mojiportal.nabunken.go.jp/>, (参照 2020-10-26).
- [9] “史料集版面ギャラリー”. <https://www.hi.u-tokyo.ac.jp/publication/dip/index.html>, (参照 2020-11-01).
- [10] “JPCOAR”. <https://schema.irdb.nii.ac.jp/ja>, (参照 2020-09-04).
- [11] 廣松毅. 人文学・社会科学分野におけるデータインフラストラクチャー構築推進事業の現状. 2020 年度統計関連学会連合大会, 2020. https://www.jsps.go.jp/j-di/data/jigyo/1_hiromatsu.pdf, (参照 2020-11-02).
- [12] “nihuiNT”. <https://int.nihui.jp/>, (参照 2020-11-02).
- [13] “ジャパンサーチ”. <https://jpsearch.go.jp/>, (参照 2020-09-04).
- [14] “Europeana”. <https://www.europeana.eu/en>, (2020-09-04).
- [15] 朝岡誠, 林正治. オープンサイエンスの取り組みとしての人文学・社会科学分野におけるデータインフラストラクチャー構築推進事業. ESTRELA, 2019, No.308, p.15-20.